The applications and social implications of data science within cardiology using machine learning.

Every year, vast numbers of lives are taken by cardiovascular diseases (CVDs). In 2019, the World Health Organisation estimates that: 17.9 million people worldwide died from CVDs including coronary heart disease (CHD), heart failure and aortic aneurysms. The identification, treatment and prevention of such illnesses are vital for sustaining cardiovascular health and saving lives. Machine learning (ML) has now, and will continue to have, a significant role in the research and management of CVDs. This essay aims to discuss the application of data science within cardiology through the use of ML as well as the societal implications of them.

ML is a component of artificial intelligence (AI), a branch of computer science where the research of machinery having human-like behaviours is the focal point, primarily using data science. Data science, historically, has been fundamental to the amelioration of society and can be defined as: *the strategy of comprehending and analysing information to conduct incisive business decisions*. In cardiology specifically, health professionals can use patient information to predict and classify CVDs to help those diagnosed with them by organising present and future treatments. ML can be divided into two major categories: supervised (SML) and unsupervised (UML) learning.

SML algorithms require multiple inputs paired with their correct corresponding output (otherwise known as a 'label'). Hence, the computer identifies a relationship between each pair, aiming to identify an overarching pattern. Once this stage reaches completion, scientists can enter new data and the computer provides a plausible prediction following the previously learnt formula.

Naive Bayes theorem (the probability of A given B, known as conditional probability) and Naive Bayes classifier (1) are SML techniques that can distribute patients with CVDs into subgroups based on patient data which may indicate which patients require prompt attention, alerting doctors to whom they should divert urgent treatment. This SML algorithm has over 80% accuracy and precision (2); thereby advancing cardiological research, enabling a brighter future for persons that struggle with CVDs.

UML are algorithms where only multiple data inputs are given - no labels are required - and the algorithm sorts them into groups called clusters. Then when fed a new input, the code, via predictive algorithm, sorts it into the cluster to which it belongs. K-means cluster is an UML technique that can be used to provide informative risk assessments of strokes and other cardiovascular events based on clinical data. Thus, allowing population health trends to be recognised and provide the capability of modifying medical care. (3)

Acknowledging risk factors and predicting how they influence the progression of CVDs in the long and short term is vital and as demonstrated by Ambale-Venkatesh and other researchers (4), ML is a fundamental asset. This study proved that traditional methods including the Cox Proportional Hazard Model (Cox-PHM) are less accurate than SML methodologies like Random Survival Forest (RSF). Random Forest (RF) is the prediction of each decision trees' prediction combined. RSF is a variation of this where the prediction can be referred to as Mortality Risk Index). 6,814 participants were sampled with various attributes, for example: age, smoking history, race, gender and the presence of diabetes. Personal information (such as heart rate, waist circumference and total cholesterol) was collected from each candidate and inputted for analysis. The results from the algorithm calculated which risk factors were the most important in terms of predicting cardiovascular events; for instance, coronary artery calcium score being the primary risk factor needed to predict CHD. Through accurately enumerating this, scientists can use this research to efficiently direct new approaches of suppressing CVD issues to protect numerous patients in the future.

Communication is a key element for humans to understand each other, accept each other and help each other; for patients to be properly treated, doctors and nurses must firstly understand them and their physical or mental disposition. Chatbots - which are ML applications that communicate with humans using natural human language - can aid this process by performing an abundance of operations ranging from remote consultations, booking appointments and communicating with clinical teams. Chatbots provide instantaneous responses to patient inquiries whilst analysing their patient history to achieve an efficient patient-specific diagnosis. Chatbots are a reliable algorithm that can provide assistance, especially when a professional is unavailable, swiftly over long distances. Other benefits can be accrued using chatbots such as decreasing care costs and decreasing emissions from transportation. IVAMED is an example of a medical chatbot that performs these services (5).

Through implementing ML algorithms, computers have the capacity to address the UK's significant issues with understaffing in hospitals. As demonstrated by the COVID pandemic, medical practices worldwide severely lack satisfactory levels of healthcare professionals, causing patient insecurity, overworked and poorly treated staff. An estimate of 490,000 jobs by 2035 are essential to satisfy hospital staff requirements (6). ML is a promising response in this context as they can ameliorate staffing and skills gaps, reducing the critical need to fill open positions. Therefore, ML allows the hospital to continue to maintain high functionality and reduce patient anxiety, whilst simultaneously lessening the mental and physical pressures on staff, as well as multiplying the number of patients treated at one time. Furthermore, ML can consistently provide accurate results whenever needed, thus ML is permanently available without threat of exhaustion or sickness. Economically, the increasing reliance of ML can, therefore, be beneficial which is an important consideration in an era of financial austerity and increasing patient demand.

However, this mechanisation of medical practises may not be wholly acceptable to the public and medical professions. Questions of sustainability (that is, the combined impacts on the environment and society as well as economics) and ethics will be increasingly important in the successful deployment of ML in medicine and treatment of CVD,

Hospitals specifically utilise large amounts of power, consequently affecting the natural world. Controlling power consumption is paramount to stabilise the environment. Due to the climate crisis, plants, animals and people alike face massive threats even up to global extinctions. Currently, we stand at a peripeteia that will ascertain whether our world can heal. However, by using more ML techniques the demand for electricity grows and so supplementary energy is demanded, which potentially will be supplied from non-renewable energy resources thus proliferating the concentration of greenhouse gases within our atmosphere and accelerating climate change. Moreover, the materials requisite for building computers ML depends on must be mined, transported and manufactured. The carbon footprint from each stage is substantial, inflicting irrevocable damage on the entire global ecosystem and endangering all that live within it. By integrating machinery unsustainably, we push the Earth and ourselves closer to the point of no return (7). Additionally, the mining of natural resources needed for these technologies has detrimental environmental impacts (8).

Furthermore, the rate of expanding fascination of industry towards the ML sector is predominantly faster than the development of appropriate controls; although ML has illustrated multiple advantages within cardiology and healthcare in general, it lacks genuinity of emotional intelligence, a cruciality within medicine. Though machines can learn from our emotions and emulate them, an imitation is merely an imitation, the warm authenticity of doctors and nurses are irreplaceable, especially to patients who suffer daily where human interaction is vital for their mental wellbeing. Moreover, if sympathy cannot be demonstrated even in healthcare environments due to its inevitable substitution by technology, what does that mean for the development of public acceptance of ML within future society as the applications of ML progress? These are questions that are not solely technological or economic but also moral and ethical in nature.

ML is also affected by racial discrimination and biases as the data used for analysis may be racially prejudiced. For example, cardiovascular events differ depending on a patients' ethnicity, thus, using data that does not account for this may lead to a misdiagnosis (7). As previously mentioned, appointments can be scheduled by ML. However, the organisation of bookings can be affected by racial prejudices. Samorani, and other professionals, concluded that black patients were approximately waiting 30% longer for a meeting compared to non-black patients (9) thus, due to racial inequity, black patients face higher risk of worsening health due to prolonged intervals between consultations. Therefore, it can be questioned whether ML should be implemented in general, as though it can enhance a portion of lives, ML can jeopardise others, obstructing the foundational principle of medical research - the research of saving all lives possible. If ML begins to choose which people to save, then what would stop people from doing it too?

Data protection and privacy is indispensable within society. The safekeeping of personal information that shields each person from possible harm is especially critical within medicine. Although, as healthcare becomes more reliant on ML, data protection issues will arise. Should ML and AI mature into an unforeseen territory, doctors are vulnerable to blame and lawsuits as well as putting patients at risk. Across industry there are already many examples of where data leakages lead to blackmail or other serious abuses because society is unprepared to control advanced technologies such as AI and ML.

Predominantly, use of data science within cardiology is being greatly changed by machine learning. By learning from patient history and data, ML can devise accurate predictions of their future wellbeing, classify different CVDs and even provide consultations. These are only a few of the various present and future applications of data science that use ML as their base. There are multiple benefits (including the production of accurate results, quick processing and meeting availability standards during times of understaffing) and consequences (data protection instability, ethical concerns and power consumption) of ML in healthcare.

Despite this, it is inarguable that technology is advancing and research into ML being implemented into cardiology will only expand. Now and in the future, machine learning will continue to play substantial roles in evolving research into cardiovascular health, ultimately enforcing society to adapt along with it.

References

(1) Berrar, D., 2018. Bayes' theorem and naive Bayes classifier. *Encyclopedia of bioinformatics and computational biology: ABC of bioinformatics*, *403*, p.412.

(2) Patil, P.B., Shastry, P.M. and Ashokumar, P.S., 2020. Machine learning based algorithm for risk prediction of cardio vascular disease (Cvd). *Journal of critical reviews*, *7*(9), pp.836-844.

(3) Flores, A.M., Schuler, A., Eberhard, A.V., Olin, J.W., Cooke, J.P., Leeper, N.J., Shah, N.H. and Ross, E.G., 2021. Unsupervised learning for automated detection of coronary artery disease subgroups. *Journal of the American Heart Association*, *10*(23), p.e021976.

(4) Ambale-Venkatesh, B., Yang, X., Wu, C.O., Liu, K., Hundley, W.G., McClelland, R., Gomes, A.S., Folsom, A.R., Shea, S., Guallar, E. and Bluemke, D.A., 2017. Cardiovascular event prediction by machine learning: the multi-ethnic study of atherosclerosis. *Circulation research*, *121*(9), pp.1092-1101.

(5) Gallent-Iglesias, D., Serantes-Raposo, S., López-Riobóo-Botana, I., Gonzalez-Vázquez, S. and Fernandez-Graña, P.M., 2023. IVAMED: Intelligent Virtual Assistant for Medical Diagnosis.

(6) Skills for Care (RTR0034), The Health Foundation, 'Summary note: workforce projections', February 2021.

(7) Gijsberts, C.M., Groenewegen, K.A., Hoefer, I.E., Eijkemans, M.J., Asselbergs, F.W., Anderson, T.J., Britton, A.R., Dekker, J.M., Engström, G., Evans, G.W. and de Graaf, J., 2015. Race/ethnic differences in the associations of the Framingham risk factors with carotid IMT and cardiovascular events. *PLoS One*, *10*(7), p.e0132321.

(7) IPCC, 2023: *Climate Change 2023: Synthesis Report.* Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115, doi: 10.59327/IPCC/AR6-9789291691647

(8) Maughan, T., The dystopian lake filled by the world's tech lust, https://www.bbc.com/future/article/20150402-the-worst-place-on-earth, 2015

(9) Samorani, M., Harris, S.L., Blount, L.G., Lu, H. and Santoro, M.A., 2022. Overbooked and overlooked: machine learning and racial bias in medical appointment scheduling. *Manufacturing & Service Operations Management*, *24*(6), pp.2825-2842.